Woche 13 Praktische Anwendungsbeispiele: Bildverarbeitung

Skript

Erarbeitet von
Dr. Jacqueline Klusik-Eckert

Lernziele **1**

Inhalt **2**

Einstieg *2*

Einsatzgebiete und Anwendungen *2*

Wie funktioniert das? *3*

menschliches Sehen vs. Computersehen *4*

Vorteile und Nachteile *6*

Black Box *7*

Take-Home Message *7*

Quellen **8**

Disclaimer **9**

Lernziele

Übersicht über verschiedene Anwendungen von KI in der Bildverarbeitung erhalten
Verständnis für die Vorteile und Limitationen von KI-basierten Lösungen im Vergleich mit traditionellen Methoden bekommen
Verständnis für ethische und datenschutzrechtliche Aspekte des Einsatzes von KI in der Bildverarbeitung gewinnen
Sensibilisiert für Biases durch Trainingsdaten werden

CC BY

Inhalt

Einstieg

Manchmal hat man so schlecht geschlafen, dass einen die Gesichtserkennung am Smartphone noch nicht mal erkennt. Kennst du auch?

Da hat die KI mal wieder einen Fehler gemacht.
Oder besser gesagt: als du die Gesichtserkennung mit deinem Gesicht vertraut gemacht hast, warst du wahrscheinlich nicht krank. Womöglich noch gut frisiert. Mit guten Lichtverhältnissen. Woher soll die KI denn wissen, dass dieses verwuschelte, unausgeschlafene etwas, das an ein Gesicht erinnert, auch du bist?

Einsatzgebiete und Anwendungen

Die Technik, die hinter dieser sehr alltäglichen Anwendung steht, kommt aus dem Bereich der automatischen Bildverarbeitung. Damit wird die Forschung über Neuronale Netze zusammengefasst, die für spezielle Einsatzbereiche des maschinellen Sehens entwickelt wurden.

Schauen wir uns doch mal ein paar Einsatzbereiche dieser KIs genauer an. Wir finden schon heute CNNs, also diese speziell trainierten KIs für das Computersehen in allen Bereichen, in denen es um das automatische Klassifizieren von Bildern geht. Gesichtserkennung für das Öffnen des Smartphones bis hin zu Überwachungsanlagen mit automatischer Erkennung von Gefahrgütern wie Schusswaffen oder Messer. Das ist hilfreich für die Sicherheitskontrolle am Flughafen. Maschinelles Sehen ist auch für autonomes Fahren wichtig, denn schließlich soll die Karre ja auf der Straße bleiben.

Verschiedene Anwendungen:

- Automatisches Klassifizieren von Bildern (Identifizieren von Objekten, Gesichtern)
- Automatisches Erkennen und Verfolgen von Bewegungen in Videodaten
- Automatisches Analysieren von medizinischen Bildern, Medizinische Bildanalyse, Medizinische Diagnostik
- Bild- und Videobearbeitung
- Überwachungskameras
- Autonomes Fahren
- Verkehrsmanagement
- …

Der Künstliche Intelligenz kommt bei der automatischen Bildverarbeitung auch dann zum Einsatz, wenn große Datenmengen schnell berechnet werden müssen. Sie ermöglichen erst die komplexen Open-World-Computerspiele mit super realistisch ausschauenden Landschaften oder atemberaubenden Special Effects in Blockbuster-Filmen.
Die verwendeten Verfahren sind auf die jeweiligen Anwendungsfelder angepasst und alle kann ich nicht erklären. Also konzentrieren wir uns mal in diesem Beispiel auf die automatische Bildverarbeitung, die für das Klassifizieren von großen Bildmengen verwendet wird.

Quelle [1, 2]

Wie funktioniert das?

Konvolutionale Neuronale Netze (CNNs) verwenden eine spezielle Art von maschinellem Lernen, um Bilder automatisch zu kategorisieren oder zu klassifizieren. Hierbei werden bestimmte Merkmale in einem Bild erkannt und dann zur Bestimmung der Kategorie verwendet.
Im ersten Schritt wird das Bild in kleine Teile unterteilt, die als Filter bezeichnet werden. Diese Filter werden über das gesamte Bild bewegt, um Merkmale wie Kanten, Farben und Formen zu erkennen.

> The first layer bases (top) and the second layer bases (bottom) learned from natural images.

Quelle [3]

Dann werden diese Merkmale kombiniert und zu höheren abstrakten Konzepten weiterverarbeitet, die für die Klassifizierung von Bedeutung sind. Hierbei werden die Ausgaben der Filter in verschiedenen Schichten verknüpft, um eine Entscheidung über die Kategorie des Bildes zu treffen.
Das CNN lernt dabei aus einer großen Anzahl vorher trainierter Bilder, um Muster und Beziehungen zwischen Merkmalen und Kategorien zu erkennen. Dieses Wissen wird dann verwendet, um neue, ungesehene Bilder automatisch zu klassifizieren.
Die ganze Forschung in diesem Bereich, also das maschinelle Sehen, wird als Computer Vision bezeichnet. Die Einsatzgebiete dieser KI-Technik sind dann wieder unglaublich vielseitig. Immer dann, wenn es um Bilder geht, und das schließt auch Bewegtbilder mit ein, haben wir es mit Computer Vision zu tun. Dabei geht es nicht nur darum, den neuronalen Netzen beizubringen, Kanten und Kontraste zu erkennen. Man zielt in diesem Teilgebiet der Informatik darauf ab, die menschliche Wahrnehmung zu simulieren und sie zu komplementieren. Die Technik kommt also auch dann zum Einsatz, wenn es um unsere Wahrnehmungsgrenzen geht: wenn etwas unglaublich schnell passieren muss oder wenn die Bildmenge einfach viel zu groß ist.

CC BY

Quelle [4, 5]

menschliches Sehen vs. Computersehen

Als gut trainierte Kunsthistorikerin kann ich mit einer Bilderflut schon ganz gut umgehen.

Einblendung von Bildern und Sortierungen

Ich habe gelernt, schnell zu sehen und nach Kategorien wie Datierung, Epoche und Stil Kunstwerke einzuordnen.

Epochen http://katrin-hemmen.blogspot.com/2012/05/time-line.html
OPEN HISTORY ARCHIVE. (o. J.). Abgerufen 2. März 2023, von https://www.openhistoryarchive.com/t-sne-map/

Bei einer überschaubaren Menge von Bildern, und genügend Zeit, bekomme ich das hin. Nach Farben sortieren, nach Größe, nach Ähnlichkeiten. Alle heraussuchen, auf denen eine Frau abgebildet ist …
Wenn ich aber jetzt aus 100.000 Bildern etwas sortieren müsste, bräuchte ich ewig. An Millionen Bilder erst gar nicht zu denken.

Dabei hat unser Sehen und das schnelle Lernen von Kategorien den KI's einiges voraus. Während eine Neuronales Netz über eine Million Bilder von Katzen braucht, um auf einem neuen, noch unbekannten Bild eine Katze zu erkennen, braucht ein Kleinkind oft nur wenige Beispiele.

https://datainnovation.org/2014/08/10000-cat-pictures-for-science/

Das Kind verknüpft mit dem Erscheinungsbild jedoch nicht nur die Form. Sehr schnell wissen Kinder, dass es sich dabei um ein Tier handelt, dass es Miau macht, dass sie flauschig sind und meistens schnell wegrennen, wenn man sie streicheln möchte. Wir Menschen sehen nicht nur einfach Formen. Wir verbinden alles Gesehene sofort mit Bedeutung, Sprache, Haptischem.

Bild menschliches vs. computersehen

Das Neuronale Netz verbindet mit einer trainierten Form eine Zeichenkette. Was diese Zeichenkette bedeutet, wird nicht verstanden.
Sie sind uns jedoch dann überlegen, wenn enorme Bildmengen in einer menschenmöglichen Gewichtigkeit verrechnet werden müssen. Sie sind besonders gut, wenn es zum Beispiel um einen Vergleich von ganzen Bildern geht, in denen man Duplikate aus großen Datenmengen herausfiltern möchte. Auch das Auffinden von einzelnen Objekten in Bildern funktioniert super. Zeige mir alle Bilder mit einer Banane.
Was mit Bananen und Katzen anfängt, kann auch zu nützlichen Szenarien hin trainiert werden. Da gibt es beispielsweise das Projekt ChextNext. Hier wurde die KI trainiert, um Krankheiten wie Tuberkulose oder ähnliches in Röntgenaufnahmen des Thorax zu erkennen. Der Algorithmus kann dabei helfen die Mediziner*innen darauf aufmerksam zu machen, wenn irgendetwas abweicht und wenn es Ungereimtheiten gibt.

Quelle [6]

Durch die Errechnung von Unterschieden können auch kleinste Abweichungen festgestellt werden, oder hättet ihr mit bloßem Auge sehen können, wo der Fehler ist?
Generell helfen diese KIs bei der Erschließung von Bilddaten, um sie zu strukturieren, zu clustern, um sie überhaupt sortierbar zu machen.

Bild von shlomz / Pixabay

Man kann sich das in etwas so vorstellen, wie ein Spürhund bei der Polizei arbeitet. Dem Hund ist es möglich, mit seiner speziellen Ausbildung für Substanzen, diese zu finden. Der Hund selbst hat aber nicht die gleiche Auffassung und Vorstellung über die Bedeutung der Substanzen. Wenn der Hund eine Ladung Drogen gefunden hat, ist ihm nicht bewusst, dass diese gegen das Betäubungsmittelgesetz verstoßen. Er weiß nur, dass er das finden soll, weil er darauf trainiert wurde. Es braucht die Kriminalkommissarin, um den Fund zu kontextualisieren.

Hund Flaticon

Das „Assistenzsystem" Hund ähnelt dem Assistenzsystem Computer Vision, das ohne Verständnis für die Gesamtsituation und den Kontext für eine relativ spezifische Aufgabe eingesetzt wird. Die Funde, bzw. Befunde müssen dann vom Menschen interpretiert und eingeordnet werden.

Quelle [7]

Vorteile und Nachteile

In allen diesen Fällen kommt die Technik zum Einsatz, wenn große Datenmengen schnell verarbeitet werden müssen. Man nutzt die Vorteile der Technik geschickt.

Höhere Geschwindigkeit bei enormer Bildmenge Größere Genauigkeit beim Erkennen von kleinsten Unterschieden Skalierbarkeit (von 2 zu Millionen und mehr)

Doch viele dieser Systeme haben direkte Auswirkungen auf den Menschen. Stellt euch doch mal vor, ihr könnt nicht in den Flieger steigen, weil eine Gesichtserkennungssoftware euch als Kriminelle erkannt haben möchte. Oder ihr bekommt eine falsche Diagnose, weil eine KI die MRT Bilder falsch ausgewertet hat.
Die Qualität der Neuronalen Netze bei der automatischen Bildverarbeitung hängt ganz stark von der Qualität der Trainingsdaten ab.
Ein Beispiel gefällig?
Neuronale Netze, die automatische Bilderkennung ermöglichen, wurden oft mit einem Datensatz trainiert: ImageNet. In diesem Datensatz wurden Bilder in einem Index von Kategorien verknüpft. Da gibt es aber nicht nur Bilder von Äpfeln, Bananen oder Katzen. Auch reale Menschen wurden in diesem Index händisch klassifiziert.

ImageNet, https://www.image-net.org/

Hier sieht man ein paar unproblematische Label wie „Trompeter” oder „Tennis Spieler”. Man findet jedoch auch eindeutig diskriminierende Kategorien, die eindeutig frauenfeindlich, rassistisch, ableistisch und altersfeindlich sind.

Klasse Ball-Buster

Quelle [8]

Ohne diese Trainingsdaten-Sets könnten wir die neuronalen Netzwerke jedoch nicht „lernen” lassen. Es ist eine schwierige, aber notwendig zu führende Diskussion, wie man damit umgeht. Generell muss man sich bewusst machen, dass hinter den Algorithmen am Ende Mathematik steht. Solche Systeme machen immer dann Fehler, wenn etwas von der Norm abweicht, sprich, wenn es auf Grund der Wahrscheinlichkeit nicht wahrscheinlich ist.

CC BY

Abhängig von guten und ausreichenden Trainingsdaten
Ungenaue Ergebnisse bei ungewöhnlichen, unbekannten Daten
Schwerwiegende Fehlentscheidungen bei Abweichungen von der „Norm“

Quelle [9]

Black Box

Dieses Problem entsteht durch die Black Box, die so eine KI auch immer sein kann. Das betrifft ethische und datenschutzrechtliche Aspekte.
Es ist zum Beispiel oft nicht sichtbar, mit welchen Daten die Gesichtserkennung trainiert wurde. Wurde immer das Recht am eigenen Bild eingehalten? Oder hat man sich illegaler Weise an Social Media Bildern bedient? Was machen eigentlich diese ganzen Selfie-Apps mit meinen Bildern? Man weiß mittlerweile, dass für ImageNet auch unzählige Privataufnahmen verwendet wurden, die über Suchmaschinen oder Dienste wie Flickr ungefragt rausgezogen wurden.

Quelle [8, 9]

Auch der Missbrauch dieser Bildverarbeitungs-KIs ist uns heute schon bekannt. In Form von Deepfakes werden sie eingesetzt, um Falschinformationen zu streuen oder andere zu verunglimpfen.
Immer wieder begegnen uns auch Ergebnisse dieser Neuronalen Netze, die anscheinend diskriminierend und vorurteilsbehaftet gegenüber Frauen, People of Colour und der LGBTQI+ Community sind.

Beth Holzer / Getty Images: AI Bias, unter https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/

Die KI selbst ist das jedoch nicht. Die hat nur gerechnet. Der Bias liegt oft schon in den Trainingsdaten. Das noch junge Feld Critical Machine Vision geht dem auf den Grund und erforscht diese Zusammenhänge

Quelle [10]

Take-Home Message

Der Einsatz von KI in der automatischen Bildverarbeitung hilft uns in vielen Forschungs- und Alltagsszenarien weiter. Trotz allem muss man sich bewusst machen, dass dieses maschinelle Sehen ganz anders funktioniert als das menschliche Sehen. Wir sind als Anwender*innen dieser Technik dafür verantwortlich, die Ergebnisse zu kontextualisieren und zu hinterfragen. Gerade in den Anwendungsbereichen, bei denen die Entscheidung der

CC BY

KI direkt eine Auswirkung auf einzelne Personen hat, muss der Mensch die letzte Entscheidungsinstanz sein.

Quelle [11]

Quellen

Quelle [1] Nischwitz, A., Fischer, M., Haberäcker, P., & Socher, G. (2019). Zusammenhang zwischen Computergrafik und Bildverarbeitung. In A. Nischwitz et al (Hrsg.), Computergrafik: Band I des Standardwerks Computergrafik und Bildverarbeitung (S. 6–24). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25384-4_2

Quelle [2] Quelle Radke, R. J. (2013). *Computer vision for visual effects*. Cambridge University Press.

Quelle [3] Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *Proceedings of the 26th Annual International Conference on Machine Learning*, 609–616. https://doi.org/10.1145/1553374.1553453

Quelle [4] Burrell, J. (2015). How the Machine „Thinks“: Understanding Opacity in Machine Learning Algorithms. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2660674

Quelle [5] Bell, P., & Offert, F. (o. J.). *Towards Open Computer Vision Methods: Visual Analytics of Machine Learning Models in the Digital Humanities* [Text]. Abgerufen 2. März 2023, von https://dh2020.adho.org/wp-content/uploads/2020/07/483_TowardsOpenComputerVisionMethodsVisualAnalyticsofMachineLearningModelsintheDigitalHumanities.html

Quelle [6] Rajpurkar, Pranav et al., Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists, in: Aziz Sheikh (Hg.), *PLOS Medicine* 15/11, 2018, e1002686.

Quelle [7] Bell, P. (2022). Computer Vision und Visualisierung als didaktische Instrumente in der Kunstgeschichte. In H. Locher & M. Männig (Hrsg.), *Lehrmedien der Kunstgeschichte: Geschichte und Perspektiven kunsthistorischer Medienpraxis* (S. 384–401). De Gruyter. https://doi.org/10.1515/9783422986251

Quelle [8] Crawford, K., & Paglen, T. (2019, September 19). *Excavating AI: The Politics of Training Sets for Machine Learning*. -. https://excavating.ai

Quelle [9] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2022). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness (arXiv:1811.12231). arXiv. http://arxiv.org/abs/1811.12231

CC BY

Quelle [10] Offert, F., & Bell, P. (2021). Perceptual bias and technical metapictures: Critical machine vision as a humanities challenge. AI & SOCIETY, 36(4), 1133–1144. https://doi.org/10.1007/s00146-020-01058-z

Quelle [11] Liu, A. (2020, März 19). Humans in the Loop: Humanities Hermeneutics and Machine Learning. DHd 2020 Keynote. https://www.youtube.com/watch?v=lnfeOUBCi3s

Disclaimer

Transkript zu dem Video „Woche 13 Praktische Anwendungsbeispiele: Bildverarbeitung“, Dr. Jacqueline Klusik-Eckert.
Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz CC-BY 4.0 veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.